TD N^o 5 : STATISTIQUES

Exercice 1 *Série discrète*

Dans une entreprise, on a relevé le nombre quotidien de bons de commande durant une période de 15 jours et on a obtenu la série suivante : 8 10 12 10 10 12 9 11 10 9 12 11 12 6 9.

1. Dépouiller ces données en complétant le tableau suivant

valeurs	
effectifs	
fréquence	
fréquence cumulée	

2. Afficher à l'écran de votre calculatrice la représentation en diagramme bâton.
3. Déterminer la moyenne et l'écart-type de cette série (résultats arrondis à 1 décimale).
4. Déterminer la médiane, les quartiles et l'intervalle inter-quartile de cette série statistique.
5. Déterminer le pourcentage de nombre de commande compris dans l'intervalle $[\overline{x} - \sigma; \overline{x} + \sigma]$

Exercice 2 *Série discrète*

Dans une entreprise, on a sélectionné 1000 employés à qui on a fait passer un test de 10 questions concernant les consignes de sécurité à connaître. L'analyse des résultats a donné le dépouillement suivant :

Nombre de réponses exactes	4	5	6	7	8	9	10
Nombre d'employés ayant obtenu cette note	48	71	262	381	167	48	23
fréquence							
fréquence cumulée							

1. Compléter le tableau ci dessus.
2. Afficher à l'écran de votre calculatrice la représentation en diagramme bâton.
3. Déterminer la moyenne $\overline{x}$ et l'écart-type σ de cette série (résultats arrondis à 2 décimale).
4. Faire une représentation des fréquences cumulées de cette série statistique.
5. Déterminer la médiane, les quartiles et l'inter-quartile de cette série statistique.
6. On estime que la formation à la sécurité a été faite correctement si au moins 95 % des employés a obtenu la moyenne (5 sur 10) ET que au moins un quart d'entre eux a obtenu au moins 8 au test. Que pouvez-vous conclure du résultat de ce test ?
7. L'année précédente, 57,8 % des employés testés avaient obtenu une note d'au moins 7 sur un test équivalent et 15,4 % s'étaient trompés à au moins la moitié des questions. Comment ces taux ont-ils évolué ? Calculer leur variation en pourcentage arrondi à 1 décimale.

Exercice 3 *Question de moyenne*

100 étudiants se sont présentés à une même épreuve de mathématiques, et la moyenne générale des notes est de 10,4. Ce groupe de 100 étudiants se divise en deux sous-groupes disjoints appelés option 1 et option 2. Sachant que la moyenne des notes des étudiants de l'option 1 est de 11 et celle des étudiants de l'option 2 est de 9,5, combien y a-t-il d'étudiants dans chaque option ?

Exercice 4 *Inconnus dans une moyenne*

Une station-service d'une grande surface a relevé, pendant une journée, la demande (en litres) de chacun de ses 890 clients de cette journée. Le tableau suivant indique le nombre de clients n ayant acheté une quantité d :

d	]5 ; 15]	]15 ; 20]	]20 ; 25]	]25 ; 30]	]30 ; 35]	]35 ; 40]	]40 ; 50]	]50 ; 60]
n	11	45	a	223	b	132	44	4

1. Sachant que la demande moyenne est $\overline{d} = 29,9$ litres, calculer les valeurs de a et b (arrondies à l'entier le plus proche).
2. Calculer alors l'écart-type s de cette série (résultat arrondi à 10^{-1} près).
3. Déterminer le pourcentage de clients dont la demande est située entre les valeurs $\overline{d} - s$ et $\overline{d} + s$.

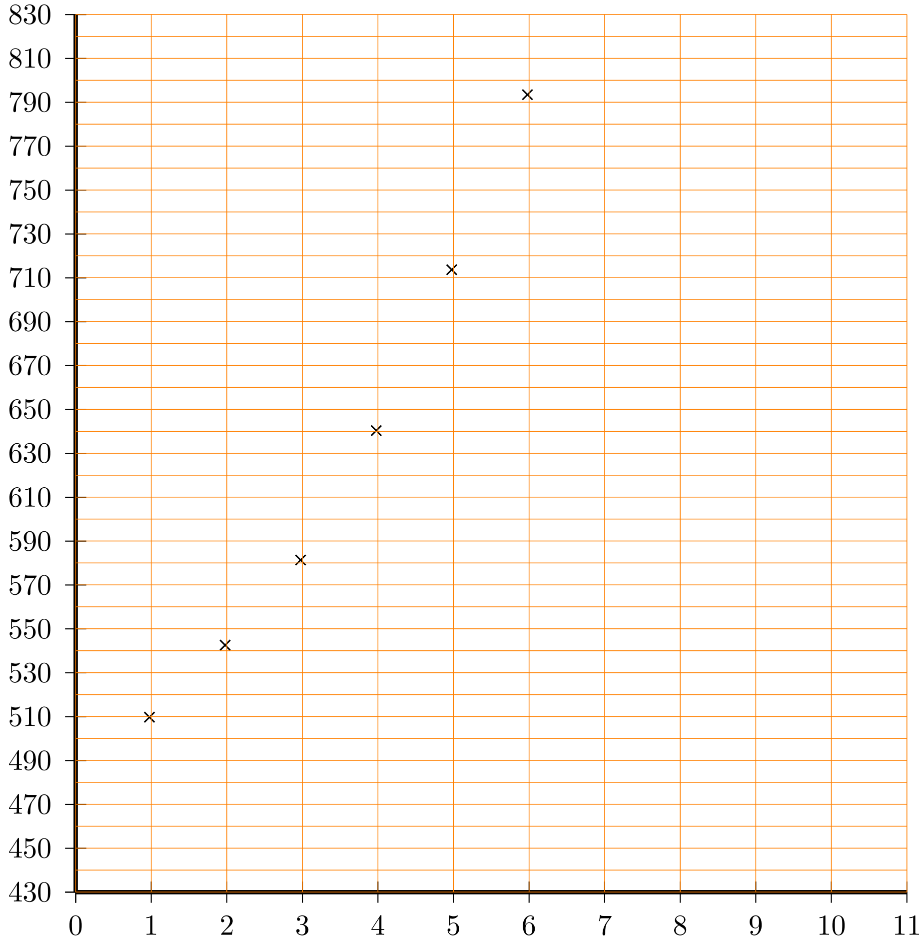
Exercice 5 *Extrapolation sur une série de deux variables*

Le tableau ci-dessous donne le montant, en milliards d'euros, des crédits accordés aux ménages entre 2001 et 2006 :

Année	2001	2002	2003	2004	2005	2006
Rang x_i	1	2	3	4	5	6
Montant y_i des crédits accordés aux ménages (en milliards d'euros)	508,9	541,8	580,5	639,5	712,9	792,7

(Source : Banque de France)

On a représenté le nuage de points de coordonnées $(x_i\ ;\ y_i)$ dans un repère orthogonal.



1. a) On appelle G le point moyen de ce nuage. Déterminer les coordonnées du point G.
   On arrondira les coordonnées du point G *au dixième.*

   b) Placer le point G sur le graphique.

2. A l'aide de la calculatrice, déterminer une équation de la droite qui réalise un ajustement affine du nuage de points de coordonnées $(x_i\ ;\ y_i)$ obtenu par la méthode des moindres carrés.

   On arrondira les coefficients au dixième.

   Dans la suite de l'exercice, on prendra comme droite d'ajustement du nuage de points de coordonnées $(x_i\ ;\ y_i)$, la droite $\mathcal{D}_1$ d'équation : $y = 57x + 430$.

3. Tracer la droite $\mathcal{D}_1$.

4. En supposant que l'ajustement affine réalisé par la droite $\mathcal{D}_1$ reste valable durant les années suivantes, déterminer à partir de quelle année le montant des crédits accordés aux ménages dépassera 980 milliards d'euros.

Exercice 6 Ajustement exponentielle

Un nom de domaine, sur Internet, est constitué de deux éléments :

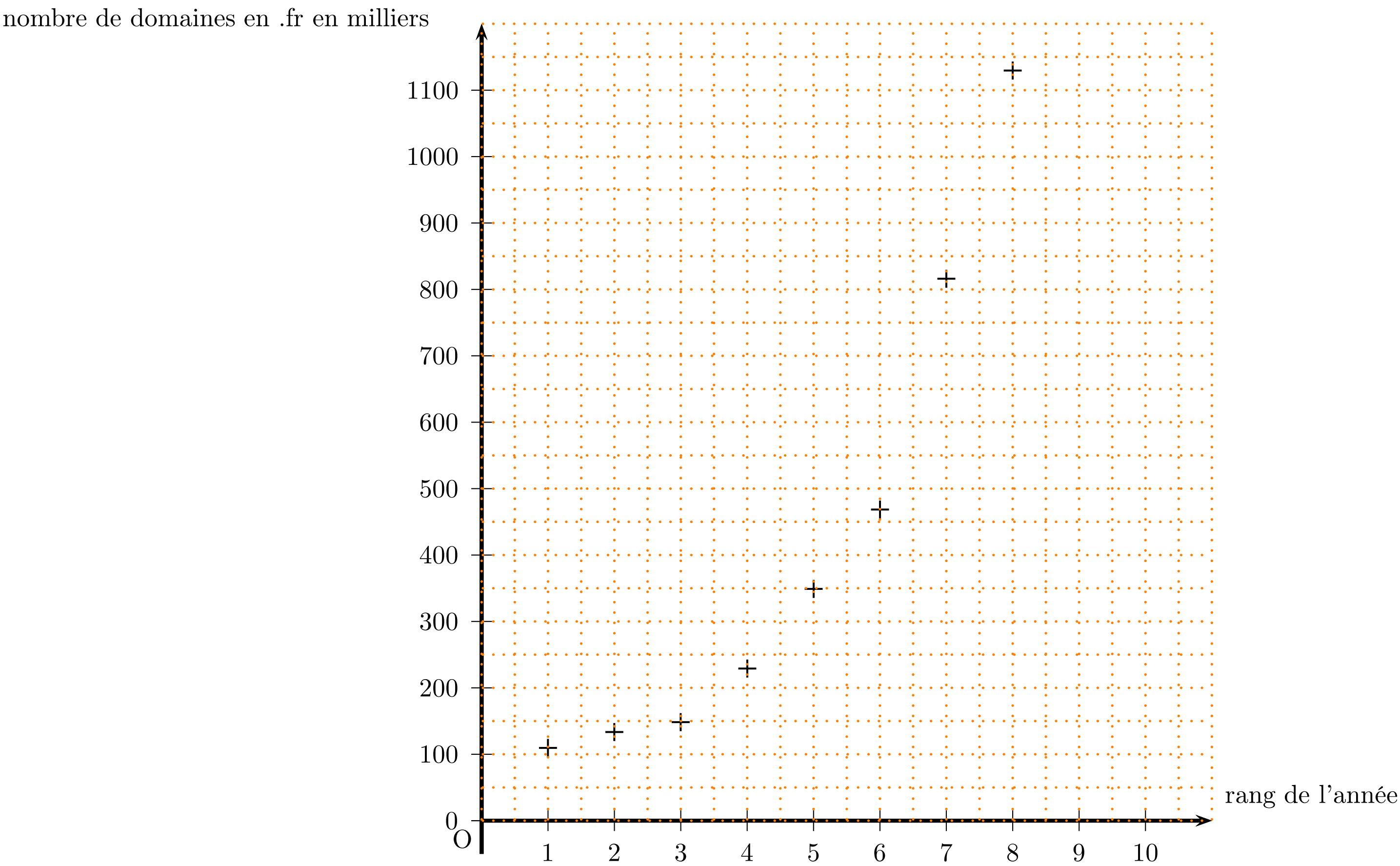
- un nom (celui d'une société, d'une marque, d'une association, d'un particulier. ..) ;
- une extension (appelée aussi suffixe) : .fr, .de, .ca, .jp, .net, .com, .org, etc.

Le tableau ci-dessous donne, en milliers, le nombre de domaines en « .fr » gérés par l'AFNIC (*Association Française pour le Nommage Internet en Coopération*), organisme qui centralise les noms de domaine Internet, pour les mois de juin des années 2001 à 2008 :

Année	2001	2002	2003	2004	2005	2006	2007	2008
Rang x_i	1	2	3	4	5	6	7	8
y_i, en milliers,	105,045	128,927	143,741	224,452	344,465	463,729	811,674	1 125,161

(Source : AFNIC, 2009)

Le nuage de points associé à cette série statistique est donné ci-dessous.



1. Calculer, en pourcentage, l'augmentation du nombre de domaines en « .fr » entre juin 2001 et juin 2002, arrondi à 1 %.

2. a) Expliquer pourquoi un ajustement affine de y en x ne semble pas justifié.

b) On cherche alors un ajustement exponentiel. Pour tout $1 \leqslant i \leqslant 8$, on pose $z_i = \ln y_i$.
Compléter le tableau ci-dessous avec les valeurs de z_i arrondies au centième :

Rang de l'année x_i	1	2	3	4	5	6	7	8
$z_i = \ln y_i$								

c) A l'aide de la calculatrice et en utilisant les données du tableau précédent, donner une équation de la droite d'ajustement de z en x par la méthode des moindres carrés sous la forme $z = ax + b$ (les coefficients seront arrondis au centième).

d) En déduire que $y = 60,95\mathrm{e}^{0,35x}$ où les coefficients sont arrondis au centième, est une ajustement exponentiel possible.

3. a) En utilisant le modèle trouvé à la question 2. d., quel est le nombre estimé de domaines en « .fr » en juin 2009 ? (Arrondir au millier).

   b) Si l'erreur commise en utilisant le modèle proposé est inférieure à 1 %, on considère que le modèle est pertinent.
   En réalité, le relevé de juin 2009 de l'AFNIC indiquait 1 412 652 domaines en « .fr ». Le modèle proposé est-il pertinent ?

4. a) Résoudre dans l'intervalle $[0\ ;\ +\infty[$ l'inéquation $60,95\mathrm{e}^{0,35x} \geqslant 10\,000$ (Arrondir au dixième).

   b) En déduire, en utilisant le modèle trouvé à la question 2. d., à partir du mois de juin de quelle année le nombre de « domaines en .fr » dépassera 10 millions.

EXERCICE 7 A SUPPRIMER

Le tableau ci-dessous donne à partir de 1998 le nombre de tués sur les routes françaises.
(Les valeurs données sont arrondies à la dizaine.)

Années	1998	1999	2000	2001	2002	2003	2004	2005	2006
Rang de l'année : x_i	0	1	2	3	4	5	6	7	8
Nombre de tués : y_i	8 440	8 030	7 640	7 720	7 240	5 800	5 590	5 320	4 700

Insee mars 2007

On donne le nuage de points $M_i\,(x_i\ ;\ y_i)$ dans un repère orthogonal.

Partie A Recherche d'un ajustement affine

1. Calculer les coordonnées du point moyen G. Placer G sur le graphique.
2. a) Déterminer à l'aide d'une calculatrice une équation de la droite d'ajustement de y en x par la méthode des moindres carres sous la forme $y = ax + b$.
   (Les valeurs de a et b seront arrondies à $0,1$ près).

   b) Tracer la droite (D) d'équation $y = -485x + 8\,660$ sur le graphique.
3. On admet que la droite (D) réalise un ajustement affine du nuage de points.
   Déterminer graphiquement une estimation du nombre de tués en 2009.
   On fera apparaître sur le graphique les traits de construction nécessaires.

Partie B Recherche d'un ajustement à l'aide d'une fonction exponentielle

On considère la fonction f définie sur l'intervalle $[0\,;\,20]$ par

$$f(x) = 8\,890\mathrm{e}^{-0,075x}.$$

1. *étude de la fonction f*

   a) Calculer la fonction dérivée f' de f sur l'intervalle $[0\,;\,20]$.

   b) Justifier que la fonction dérivée f' est strictement négative sur l'intervalle $[0\,;\,20]$.

   c) En déduire le sens de variations de la fonction f sur l'intervalle $[0\,;\,20]$.

d) Dresser le tableau de variation de la fonction f sur l'intervalle $[0\,;\,20]$.

2. *Représentation de la fonction f*

a) Recopier et compléter le tableau de valeurs suivant ; on donnera les valeurs approchées entières arrondies à la dizaine la plus proche.

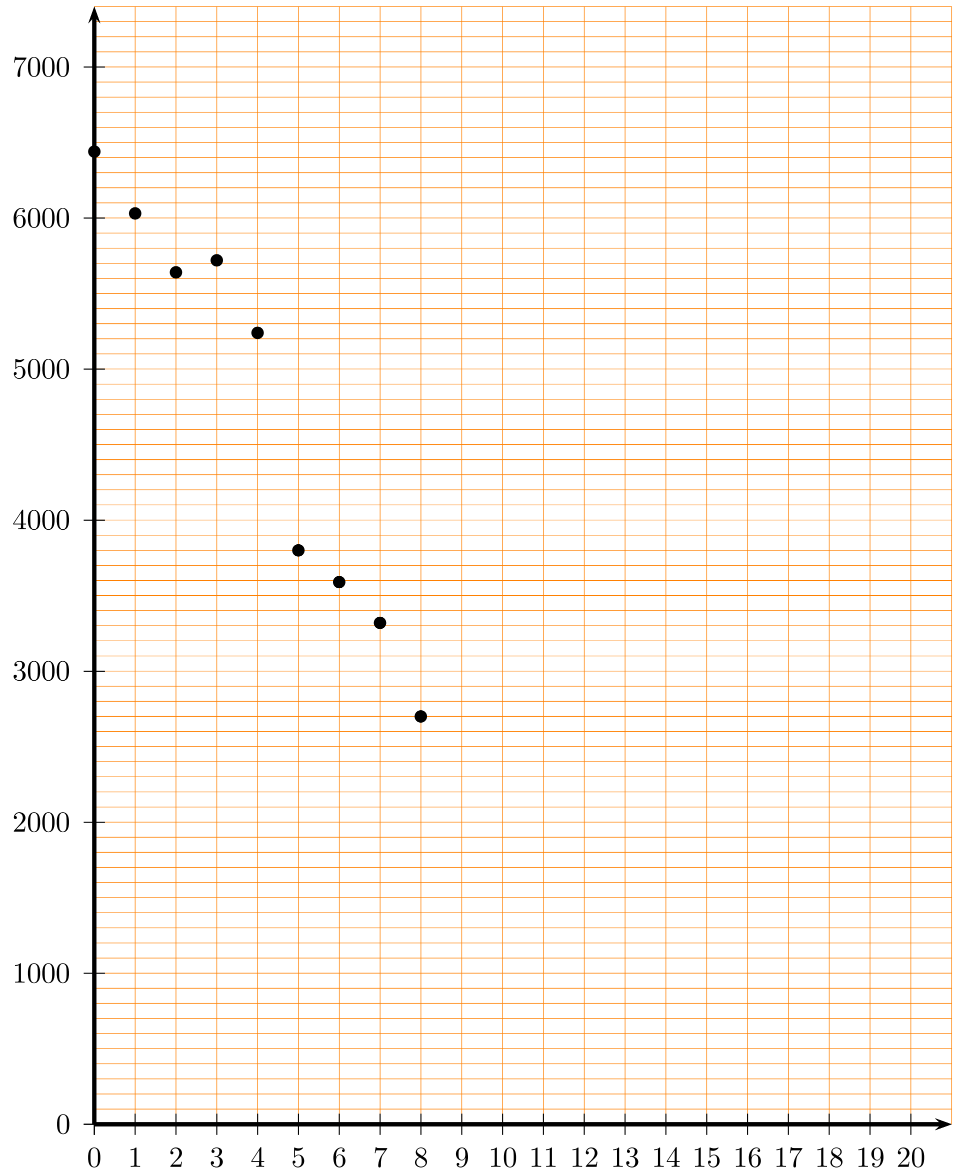
x	0	2	4	6	8	10	12	14	16	18	20
$f(x)$		7 650	6 590		4 880			3 110			

b) En utilisant les valeurs du tableau de la question précédente, construire la courbe représentative de la fonction f sur le graphique.

c) Prouver que la fonction f réalise un deuxième ajustement du nuage de points. Il sera nécessaire de construire un raisonnement en plusieurs étapes, notamment en introduisant la variables $z = \ln y$.
Estimer par la méthode de son choix le nombre de tués en 2009.
On fera apparaître sur le graphique les traits de construction nécessaires.

Partie C Comparaison des deux ajustements

1. A l'aide de l'ajustement affine de la partie A, estimer, par un calcul, en quelle année le nombre de tués sera inférieur à 2 500.
2. A l'aide de l'ajustement de la partie B, estimer, par un calcul, en quelle année le nombre de tués sera inférieur à 2 500.
3. Quel est, parmi les deux ajustements étudiés, celui qui semble le plus réaliste ? Expliquer son choix.



Exercice 7

Le tableau ci-dessous donne les cumuls des nombres d'entrées de cinq films sortis au cours de l'année 2006, d'une part en région parisienne, d'autre part sur la France dans son ensemble. (*source : « le film français », chiffres arrêtés au* 3 *avril* 2007)

Film	Indice i ($1 \leqslant i \leqslant 5$)	Nombres d'entrées en région parisienne en centaines de milliers : x_i	Nombres d'entrées en France en centaines de milliers : y_i
Pirates des Caraïbes 2	1	10	75
Arthur et les Minimoys	2	9	62
Da Vinci Code	3	7,5	41,5
Ne le dis à personne	4	6,5	32
Indigènes	5	5	29,5

1. a) Représenter le nuage de points associé à la série statistique $(x_i\ ;\ y_i)$ ($1 \leqslant i \leqslant 5$) dans le plan rapporté à un repère orthogonal (unités graphiques : 1 cm pour une centaine de milliers d'entrées sur l'axe des abscisses et 1 cm pour 10 centaines de milliers d'entrées sur l'axe des ordonnées).

   b) Déterminer les coordonnées du point moyen G de cette série et placer G dans le repère précédent.

   c) Donner, à l'aide de la calculatrice, une équation de Δ, droite d'ajustement de y en x obtenue par la méthode des moindres carrés (les coefficients sont arrondis au dixième). Tracer cette droite dans le repère précédent.

   d) En utilisant cette approximation affine, calculer le nombre d'entrées cumulées sur la France qu'on aurait pu prévoir pour le film « Les bronzés 3 » sachant qu'il en a réalisé 1 140 000 en région parisienne (on arrondira le résultat à la dizaine de milliers d'entrées).

2. La forme du nuage de points ci-dessus suggère de faire un ajustement par une courbe de type exponentiel d'équation $y = A\mathrm{e}^{Bx}$ (où A et B sont des réels). Pour cela on pose d'abord $z = \ln(y)$.

   a) Recopier et compléter le tableau suivant avec des valeurs de z_i arrondies à 10^{-2} ($1 \leqslant i \leqslant 5$).

x_i	10	9	7,5	6,5	5
y_i	75	62	41,5	32	29,5
$z_i = \ln(y_i)$					

   b) Déterminer, à l'aide de la calculatrice, une équation de la droite d'ajustement de z en x par la méthode des moindres carrés (*les coefficients seront arrondis au millième*).

   c) En utilisant la relation $z = \ln(y)$ déterminer alors les valeurs arrondies à 10^{-3} des réels A et B tels que $y = A\mathrm{e}^{Bx}$.

   d) En utilisant l'approximation $y \approx 9,689\mathrm{e}^{0,202x}$, quel nombre d'entrées, cumulées sur la France aurait-on pu prévoir pour le film « Les bronzés 3 » sachant qu'il en a réalisé 1 140 000 en région parisienne ? On arrondira le résultat au millier d'entrées.

3. Le nombre d'entrées en fin d'exploitation pour ce film sur la France a été de 10 300 000.
   Lequel des deux ajustements semble le plus approprié ?

Exercice 7 *Multiples ajustements* A FAIRE APRES CHAP ln

On rappelle que :

- Le taux d'emploi d'une classe d'individus est calculé en rapportant le nombre d'individus de la classe ayant un emploi au nombre total d'individus dans la classe.
- Un individu âgé de 55 ans à 64 ans est appelé un « senior ».
- UE désigne l'Union européenne.

Selon un rapport de l'INSEE :
« Le taux d'emploi des personnes âgées de 55 à 64 ans est considéré comme un levier privilégié pour limiter l'exclusion de ces personnes du marché du travail et maîtriser les dépenses de retraites.
En 2008, il est de 45,6 % dans l'UE, mais seulement de 38,3 % en France alors que l'objectif de l'UE comme de la France est d'atteindre 50 % en 2010. »

Le but de l'exercice est de vérifier si la France a atteint l'objectif visé par l'UE.
Dans tout l'exercice, le taux d'emploi sera exprimé en pourcentage. Les valeurs approchées seront arrondies au dixième.

Partie A Etude statistique et interpolation de données

Le tableau ci-dessous indique le taux d'emploi des seniors en France entre 1992 et 1998 :

Année	1992	1993	1994	1995	1996	1997	1998
Rang de l'année x_i	0	1	2	3	4	5	6
Taux d'emploi des seniors en % y_i	29,8	29,7	29,6	29,6	29,4	29	28,3

Source : INSEE, Eurostat

1. Déterminer, en utilisant la calculatrice, l'équation de la droite d'ajustement affine de y en x obtenue par la méthode des moindres carrés.
2. Selon cet ajustement, déterminer le taux d'emploi des seniors en 1999.
3. Selon cet ajustement, déterminer si la France a atteint l'objectif fixé en 2010.

Partie B Interpolation de données à l'aide d'un second modèle

Le taux d'emploi des séniors en France est en réalité de 28,8 % en 1999 et on admet qu'à partir de l'année $2000 + n$, il est donné par l'expression $29,9 \times 1,037^n$ où n désigne un entier naturel. Selon ce modèle, déterminer :

1. Le taux d'emploi des seniors en 2010.
2. A partir de quelle année, la France aura atteint son objectif.

Partie C Extrapolation de données selon un troisième modèle

Le tableau ci-dessous indique le taux d'emploi des seniors en France entre 2001 et 2009 :

Année	2001	2002	2003	2004	2005	2006	2007	2008	2009
Rang de l'année x_i	9	10	11	12	13	14	15	16	17
Taux d'emploi des seniors en % y_i	31,9	34,7	37	37,8	38,5	38,1	38,2	38,2	38,9

Source : INSEE, Eurostat

Désormais, à partir de 2001, on choisit un modèle logarithmique et on admettra qu'à partir de 2001, le taux d'emploi des seniors est donné par la fonction f définie sur $[9 ; +\infty[$ par

$$f(x) = a\ln(x+1) + b \text{ où } a \text{ et } b \text{ désignent deux nombres réels.}$$

1. En considérant les années 2001 et 2006, écrire le système d'équations que doivent vérifier a et b.
2. En déduire que $a = \dfrac{6,2}{\ln 1,5}$.

   Dans la suite, on admettra que $a = 15,3$ et $b = -3,3$.
3. Selon ce modèle, déterminer à partir de quelle année, la France aura atteint son objectif.